

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:58 ; Search time 91.75 Seconds
(without alignments)
21.119 Million cell updates/sec

Title: US-09-331-631a-21_COPY_32_91

Perfect score: 343
Sequence: 1 TENPCAGRCLOSCQOEPPDDL.....DTGATNORHPGERTGRGP 60

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	343	100.0	614	1 AH11_ARAHY	P43237 arachis hyp
2	318	92.7	626	1 AH12_ARAHY	P43238 arachis hyp
3	98.5	28.7	639	1 GLCA_SOTBN	P11827 glycine max
4	93	27.1	605	1 GLCA_SOTBN	P13916 glycine max
5	66.5	19.4	771	1 SM3A_HUMAN	O14563 homo sapien
6	66	19.2	605	1 VCLA_GOSHI	P09799 gossypium h
7	64	18.7	1300	1 IRR_CAVPO	P14617 cavia porce
8	63.5	18.5	588	1 VCLB_GOSHI	P09801 gossypium h
9	63.5	18.5	772	1 SM3A_MOUSE	O08685 mus musculu
10	63.5	18.5	772	1 SM3A_MOUSE	O08685 mus musculu
11	63	18.4	591	1 UL49_EBV	P63548 ratius norv
12	61	17.8	545	1 A4C1_SOLTU	P14347 epstein-bar
13	61	17.8	545	1 A4C1_SOLTU	P14347 epstein-bar
14	60.5	17.6	777	1 SM3D_HUMAN	P31685 solanum tub
15	60.5	17.6	777	1 SM3D_HUMAN	P31685 solanum tub
16	60.5	17.6	777	1 SM3D_HUMAN	P31685 solanum tub
17	59.5	17.3	419	1 VEGC_HUMAN	O95025 homo sapien
18	59.5	17.3	772	1 SM3A_CHICK	P98160 homo sapien
19	59	17.2	147	1 IP2A_SOLTU	O90607 gallus gall
20	59	17.2	356	1 TRBM_BOVIN	O41435 solanum tub
21	59	17.2	448	1 FBLS_MOUSE	P06579 bos taurus
22	59	17.2	448	1 FBLS_MOUSE	P06579 bos taurus
23	58.5	17.1	751	1 SM3C_CHICK	O9WV88 ratius norv
24	58.5	17.1	106	1 THG1_NICPA	O42236 gallus gall
25	58.5	17.1	785	1 SZLB_BRARE	O24115 nicotiana p
26	58.5	17.1	785	1 SM3E_CHICK	O9W686 brachydanio
27	58.5	17.1	4753	1 LRP_CABEL	O42267 gallus gall
28	58	16.9	448	1 FBLS_HUMAN	O04833 caenorhabdi
29	57.5	16.8	487	1 VSMB_TRYBR	O9UBX5 homo sapien
30	57.5	16.8	707	1 BMP1_XENLA	P20947 trypanosoma
31	57	16.6	2768	1 THYG_HUMAN	P98070 xenopus lae
32	57	16.6	88	1 OM3_CHLTR	P21355 chlamydia t
33	57	16.6	153	1 IP2K_SOLTU	P01080 solanum tub
			751	1 SM3C_HUMAN	O99985 homo sapien

34	57	16.6	751	1 SM3C_MOUSE	O62181 mus musculu
35	57	16.6	823	1 CN8A_MOUSE	O88502 mus musculu
36	56	16.3	147	1 IP2Y_SOLTU	O41489 solanum tub
37	56	16.3	1382	1 INSR_HUMAN	P06213 homo sapien
38	56	16.3	1382	1 INSR_HUMAN	P06213 homo sapien
39	56	16.3	2813	1 VWF_CANFA	O28295 canis famli
40	55.5	16.2	4344	1 LRP1_HUMAN	O07934 homo sapien
41	55.5	16.2	562	1 ATP2_HEYBR	P29685 hevea brasl
42	55.5	16.2	730	1 BMP1_HUMAN	P13497 homo sapien
43	55.5	16.2	761	1 SM3D_CHICK	O90663 gallus gall
44	55.5	16.2	991	1 BMP1_MOUSE	P98063 mus musculu
45	55.5	16.2	1106	1 STC_DROME	P40798 drosophila
			2476	1 ZAN_PIG	O28983 sus scrofa

ALIGNMENTS

RESULT 1	
AC P43237:	STANDARD; PRT; 614 AA.
DT 01-NOV-1995 (Rel. 32, Last sequence update)	
DT 01-NOV-1997 (Rel. 35, Last annotation update)	
DE ALLERGEN ARA H 1, CLONE P17 (ARA H I).	
OS Arachis hypogaea (Peanut).	
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; kosisidae; eurosoids I; Fabales; Fabaceae; Papilionoideae; Arachis.	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN-CV. FLORUNNER;	
RX MEDLINE-96013631; Pubmed-7560062;	
RA Burke A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;	
RT "recombinant peanut allergen Ara h I expression and IgE binding in patients with peanut hypersensitivity."	
RL J. Clin. Invest. 96:1715-1721(1995).	
CC -1- SIMILARITY: TO 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLYCININ, ETC.).	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC EMBL: L38853; AAA60336.1; -.	
CC HSSP: P50477; ICAM.	
DR INTERPRO: IPR001113; -.	
DR PRAM: PF00546; Seedstore_7s; 1.	
KW Allergen.	
SQ SEQUENCE 614 AA; 70283 MW; 1DDACF217EEC5F31 CRC64;	
Query Match 100.0%; Score 343; DB 1; Length 614;	
Best Local Similarity 100.0%; Pred. No. 6.4e-31;	
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY 1 TENPCAGRCLOSCQOEPPDLKAKACSRCTKLEYDPVCYDGTATNORHPGERTGRGP 60	
Db 32 TENPCAGRCLOSCQOEPPDLKAKACSRCTKLEYDPVCYDGTATNORHPGERTGRGP 91	
RESULT 2	
ID AH12_ARAHY	STANDARD; PRT; 626 AA.
AC P43238:	
DT 01-NOV-1995 (Rel. 32, Last sequence update)	
DT 01-NOV-1997 (Rel. 35, Last annotation update)	
DE ALLERGEN ARA H 1, CLONE P41B (ARA H I).	

```

OC Arachis hypogaea (Peanut) .
OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Arachis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. FLORUNNER.
RX MEDLINE=96013631; PubMed=7560062;
RA Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;
RT "Recombinant peanut allergen Ara h I expression and IgE binding in
RL patients with peanut hypersensitivity."
RJ J. Clin. Invest. 96:1715-1721(1995).
CC -I- SIMILARITY: TO 7S SEED STORAGE PROTEINS (PHASOLIN, VICLIN,
CC CONVULCIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: L34402; AAB0861.1; -.
DR HSSP: P50477; ICAM.
DR INTERPRO: IPR001113; -.
DR Pfam: PF00546; Seedstore_7s; 1.
RW Allergen.
SO SEQUENCE 626 AA; 71345 MW; 1A6BBBE419DDE3 CRC64;

Query Match          92.7%; Score 318; DB 1; Length 626;
Best Local Similarity 90.6%; Pred. No. 3.9e-28;
Matches 58; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY      1 TENPAORCLGCGQOEPPDLKKAECESRCTKLEYPCRYD---TGATNORHPGEPTR 56
Db      34 TENPAORCLGCGQOEPPDLKKAECESCTKLEYPCRYDPRGHGTITTONSPGEERT 93
        |||||
OY      57 GRQP 60
        ||||
Db      94 GRQP 97

RESULT 3
AC GLCX_SOYBN STANDARD: PRT; 639 AA.
ID ID GLCX_SOYBN STANDARD: PRT; 639 AA.
AC P11827;
DT DT 01-OCT-1989 (Rel. 12, Created)
DT DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE DE BETA-CONGLYCININ, ALPHA' CHAIN PRECURSOR.
GN GN CG-1.
OS Glycine max (Soybean) .
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. FLORUNNER.
RX MEDLINE=86250867; PubMed=3013879;
RA Doyle J.J., Schuler M.A., Godette W.D., Zenger V., Beachy R.N.,
RA Slightom J.L.;
RT "The glycosylated seed storage proteins of Glycine max and Phaseolus
RT vulgaris. Structural homologies of genes and proteins..";
RJ J. Biol. Chem. 261:9228-9238(1986).
RN [2]
RP SEQUENCE OF 340-639 FROM N.A.
RC MEDLINE=83143289; PubMed=6897678;
RA Schuler M.A., Ladin B.F., Pollaco J.C., Freyer G., Beachy R.N.;
RT "Structural sequences are conserved in the genes coding for the
RT alpha, alpha' and beta-subunits of the soybean 7S seed storage
RT protein..";
RL Nucleic Acids Res. 10:8245-8261(1982).
```

```

CC -1- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC -1- SUBUNIT: THE ALPHA-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
CC VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVITICILIN, CONGLYCININ, ETC.).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: M13759; AAB01374.1; -.
CC DR EMBL: J01290; NOT_ANNOTATED_CDS.
CC DR PIR: B24810; B24810.
CC DR HSSP: P02853; 2PML.
CC DR INTERPRO: IPR001113; -.
CC DR PFAM: PF00546; Seedstore_7s; 2.
CC KW Seed storage protein; signal; Glycoprotein; Multigene family.
CC FT SIGNAL 1 22
CC FT PROPEL 23 62
CC FT CHAIN 63 639 BETA-CONGLYCININ, ALPHA' CHAIN.
CC FT CARBOHD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHD 551 551 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CONFLICT 543 543 P -> L (IN REF. 2).
CC FT CONFLICT 549 549 M -> V (IN REF. 2).
CC FT CONFLICT 608 608 S -> T (IN REF. 2).
CC SQ SEQUENCE 639 AA; 74325 MW; 469BF24C79651E3F CRC64;
CC -----
CC Query Match 28.7%; Score 98.5; DB 1; Length 639;
CC Best Local Similarity 30.6%; Pred. No. 0.00083;
CC Matches 22; Conservative 10; Mismatches 25; Indels 15; Gaps 3;
CC -----
QY 2 ENPCARCIOSQOGPPDLKOKACSRCT--KLEVDPCRVDTG-----ATNQR 48
DB 31 QNPSPNKKCLIRSNSEKSDYRNQACHARCMLKVEEBEC--EGQILPPRPHPEREROO 88
QY 49 HPPGERTGRGP 60
DB 89 HGEKEDEGECP 100
CC -----
RESULT 4
AC GLCA_SOYBN STANDARD; PRT; 605 AA.
AC P13916;
AC 01-JAN-1990 (Rel. 13, Created)
AC 01-APR-1990 (Rel. 14, Last sequence update)
AC 15-DEC-1998 (Rel. 37, Last annotation update)
AC BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.
AC Glycine max (Soybean).
AC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
AC Fabales; fabaceae; Papilionoideae; Glycine.
AC [1]
AC SEQUENCE FROM N.A.
AC TISSUE=COYLEDON;
AC MEDLINE=91355860; Pubmed=2103438;
AC Sebastiani F.L., Farrell L.B., Schuler M.A., Beachy R.N.;
AC "Complete sequence of a cDNA of alpha subunit of soybean beta-
AC conglycinin."
AC Plant Mol. Biol. 15:197-201(1990).
AC -1- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
AC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
AC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
AC -1- SUBUNIT: THE ALPHA-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN

```

```
CC      CC      VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC      CC      -I- SUBCELLULAR LOCATION: EMBRO AXIS, AND COTYLEDONARY MEMBRANE-BOUND
CC      CC      VACUOLAR PROTEIN BODIES.
CC      CC      -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC      CC      CONVULGIN, CONVOLYCININ, ETC.).
CC      CC      -----
CC      CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      CC      use by non-profit institutions as long as its content is in no way
CC      CC      modified and this statement is not removed. Usage by and for commercial
CC      CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      CC      or send an email to license@isb-sib.ch).
CC      CC      -----
DR      DR      EMBL; X17698; CAAB3691.1; -.
DR      DR      PIR; S14681; FMSYBA.
DR      DR      HSSP; P50477; ICAM.
DR      DR      INTERPRO: IPR001113; -.
DR      DR      PFAM; PF00546; Seedstore_7s; 1.
KW      KW      Seed storage protein; Signal; Glycoprotein; Multigene family.
FT      FT      SIGNAL                     1         22
FT      FT      PROPEP                    23         62
FT      FT      CHAIN                      63        605
FT      FT      CARBOHYD                     261       261
FT      FT      CARBOHYD                     517       517
SO      SO      SEQUENCE                  605 AA; 70293 MW; CBDBA50506BBC57 CRC64;

Query Match                               27.1%; Score 93; DB 1; Length 605;
Best Local Similarity 31.4%; Pred. NO. 0.0032;
Matches 22; Conservative 8; Mismatches 16; Indels 24; Gaps 3.

QY      QY      2 ENPCAGRCIGSCQGPDDLKOKACESRCTKLEYD-----PCRYVDGATNORHP 50
Db      Db      31 ENPKHKKCTGCSNSRDSRYNQACHARCNLKVEKECEFEGETPPRR-----PPQHP 83
          ||| :||| | | : || || | : ||
QY      QY      51 -----PGER 54
Db      Db      84 EREPOQGEEK 93
          ||| :

RESULT   5
SM3A_HUMAN STANDARD; PRT; 771 AA.
ID      ID      SM3A_HUMAN STANDARD; PRT; 771 AA.
AC      AC      Q14563;
DT      DT      30-MAY-2000 (Rel. 39, Created)
DT      DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      DE      SEMAPHORIN 3A PRECURSOR (SEMAPHORIN III) (SEMA III),
GN      GN      SEMA3A.
OS      OS      Homo sapiens (Human).
OC      OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      RN      [1]
RP      RP      SEQUENCE FROM N.A.
RC      RC      TISSUE=FETAL BRAIN;
RX      RX      MEDLINE=94094332; PubMed=8269517;
RA      RA      Kolodkin A.L., Matthes D.J., Goodman C.S.;
RT      RT      "The semaphorin genes encode a family of transmembrane and secreted
RL      RL      growth cone guidance molecules."
LN      LN      Cell 75:1389-1399(1993).
RN      RN      [2]
RP      RP      SEQUENCE OF 1-37 FROM N.A.
RA      RA      Moessner J., Minx P., Hinds K., Stromwalt C.;
RL      RL      submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN      RN      [3]
RP      RP      SEQUENCE OF 39-182 FROM N.A.
RA      RA      Rohlfing T., Tin-Wollam A.M., Duckels G.;
RL      RL      submitted (JUN-1998) to the EMBL/Genbank/DDBJ databases.
CC      CC      -I- FUNCTION: INDUCES THE COLLAPSE AND PARALYSIS OF NEURONAL GROWTH
CC      CC      CONES. COULD SERVE AS A LIGAND THAT GUIDES SPECIFIC GROWTH CONES
CC      CC      BY A MOTILITY-INHIBITING MECHANISM. BINDS TO THE COMPLEX
CC      CC      NEUROFILIN-1/PLEXIN-1 (BY SIMILARITY).
```

```

CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- DOMAIN: STRONG BINDING TO NEUROPHILIN IS MEDIATED BY THE CARBOXY
CC THIRD OF THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L26081; AAA65938.1; -
CC EMBL; AC0004451; AAC60185.1; -
CC EMBL; AC0004848; AAC78622.1; -
CC MIM; 603961; -
CC INTERPRO; IPR001627; -
CC INTERPRO; IPR003006; -
CC PFM; PF01403; Sema; 1.
CC PFM; PF00047; Ig; 1.
CC KW Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
CC Developmental protein; Glycoprotein.
CC SIGNAL 1 20
CC CHAIN 21 771
CC DOMAIN 240 538
CC DOMAIN 642 729
CC DOMAIN 727 769
CC DISULFID 649 722
CC CARBOHYD 53
CC CARBOHYD 125 125
CC CARBOHYD 590 590
CC SEQUENCE 771 AA; 88889 MW; 9985FBD3EAD8456 CRC64;
CC -----
QY Query Match 19.4%, Score 66.5; DB 1; Length 771;
Best Local Similarity 35.2%; Pred. No. 3.5;
Matches 19; Conservative 5; Mismatches 23; Indels 7; Gaps 2;
DB 6 AQRCLSGQGPDDLKQKACESRCIKLEYDPCVYDTGATNQRHPGERTRRQ 59
510 AQLPLHRC-----DIYGRACAECC--IARDPYCAMDGSACSRYPFPTARRRRRQ 556
RESULT 6
VCLA_GOSHI ID STANDARD; PRT; 605 AA.
AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1998 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chuan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOLEONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASOLIN, VICILIN,
CC CONVITICLIN, CONGLICININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```
CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19378; AAA3069.1; -.
DR PIR; S06398; A06398.
DR HSSP; P50477; ICAX.
DR INTERPRO; IPR00113; -.
DR PFAM; PF00546; Seedstore_7s; 1.
KW Seed storage protein, Signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 605 AA; 71049 MW; C9DB39371C976953B CRC64;

Oy 9 CLASC-----QGEPLDLKQ-KACESRCRKLEDP----RCVYDTGATNQRHP-PGER 54
Db 102 CGSOSREQYEKGQQQQP--KQFKCGQRQMQRPERPKQGVCKEGRYOEDPMRGER 159

RESULT 7
ID IRR_CAVPO STANDARD; PRT; 1300 AA.
AC P14617;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE INSULIN RECEPTOR-RELATED PROTEIN PRECURSOR (EC 2.7.1.112) (IRR)
DE (IR-RELATED RECEPTOR).
DE INSR.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Cavillidae; Cavia.
OC [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89359245; PubMed=2768234;
RA Shier P., Watt V.M.;
RT "Primary structure of a putative receptor for a ligand of the Insulin family."
RL J. Biol. Chem. 264:14605-14608(1989).
CC -I- FUNCTION: THIS RECEPTOR PROBABLY BINDS AN INSULIN RELATED PROTEIN AND HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
CC -I- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN TYROSINE PHOSPHATE.
CC -I- SUBUNIT: PROBABLE Tetramer OF 2 ALPHA AND 2 BETA CHAINS LINKED BY DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAIN CARRY THE KINASE DOMAIN.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-PROTEIN KINASES.
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05047; AAA37044.1; -.
DR PIR; A36502; A36502.
DR HSSP; P06213; IIRK.
DR INTERPRO; IPR000494; -.
DR INTERPRO; IPR000719; -.
DR INTERPRO; IPR001245; -.
DR INTERPRO; IPR001777; -.

```

[illegible]


```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR: TISSUE-BRAIN:
RX MEDLINE-97073089; PubMed-8915837;
RA Giger R.J., Wolfer D.P., De Wit G.M.J., Verhaagen J.;
RT "Anatomy of rat semaphorin III/collapsin-1 mRNA expression and
RT relationship to developing nerve tracts during neuroembryogenesis.";
RL J. Comp. Neurol. 375:378-392(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN GUIDING GROWING AXONS TOWARDS THEIR
CC TARGETS BY FORMING A MOLECULAR BOUNDARY THAT INSTRUCTS AXONS TO
CC ENGAGE IN THE FORMATION OF SPECIFIC NERVE TRACTS. BINDS TO
CC NEUROFILIN.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- DEVELOPMENTAL STAGE: AT E-11, EXPRESSION WAS RESTRICTED TO THE
CC OLFACTORY PIT, THE BASAL AND ROSTRAL SURFACE OF THE TELENEPHALIC
CC VESICLE, THE EYE ANLAGE, THE EPITHELIUM OF RATHKE'S POUCH, AND
CC SOMITES. AT LATER DEVELOPMENTAL STAGES, IT WAS WIDELY DISTRIBUTED
CC IN NEURONAL AS WELL AS IN MESENCHYMAL AND EPITHELIAL STRUCTURES
CC OUTSIDE THE NERVOUS SYSTEM. AFTER BIRTH, MESENCHYMAL LEVELS
CC DECREASED RAPIDLY AND EXPRESSION BECAME RESTRICTED TO SPECIFIC
CC SETS OF NEURONS IN THE CNS. IN THE MATURE CNS, IT IS DETECTABLE IN
CC MITRAL CELLS, NEURONS OF THE ACCESSORY OLIVAR AND CEREBRAL CORTEX,
CC CEREBELLAR PURKINJE CELLS, AS WELL AS A SUBSET OF CRANIAL AND
CC SPINAL MOTONEURONS.
CC -1- DOMAIN: STRONG BINDING TO NEUROFILIN IS MEDIATED BY THE CARBOXY
CC THIRD OF THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X95286; CAA64607.1; -.
DR INTERPRO: IPR001627; -.
DR INTERPRO: IPR003006; -.
DR PRAM: PF01403; Sema: 1.
DR PFAM: PF00047; Ig; 1.
KM Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
KM Developmental protein; Glycoprotein.
FT SIGNAL 1 20
FT CHAIN 1 772
FT DOMAIN 240 538
FT DOMAIN 643 730
FT DOMAIN 728 770
FT DISULFID 650 723
FT CARBOHYD 53 53
FT CARBOHYD 125 125
FT CARBOHYD 591 591
FT SEQUENCE 772 AA; 88808 MW; 240907812FF9F2D2 CRC64;

```

```

Query Match      18.5%; Score 63.5; DB 1; Length 772;
Best Local Similarity 33.3%; Pred. No. 7.5;
Matches 18; Conservative 6; Mismatches 23; Indels 7; Gaps 2;

```

```

QY 6 AGRCLSCQOEPPDLKOKACESEKCTKLEYDPCVYDGTATNORHPGEPTGRGQ 59
DB 510 AQLPLHNC-----DIYKACAEC--LARDPYCAMDSSCSKRYFTAKRRTRRQ 556

```

```

RESULT 11
ID UL49_EBV
AC P14347;
DT 01-JAN-1990 (Rel. 13, Created)
STANDARD; PRT; 591 AA.

```

```

DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PROTEIN BFRF2.
GN BFRF2.
OS Epstein-Barr virus (strain 95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-94270667; PubMed-6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Telford P.S., Watford B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BFRF2,
CC HVS-1 66, AND HCMV UL49.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: V01555; CAA24880.1; ALT_INIT.
DR KM Early protein.
SQ SEQUENCE 591 AA; 63977 MW; EE63F1E0721912E CRC64;

```

```

Query Match      18.4%; Score 63; DB 1; Length 591;
Best Local Similarity 34.4%; Pred. No. 6.6;
Matches 22; Conservative 5; Mismatches 19; Indels 18; Gaps 5;

```

```

QY 5 CAQRCLOSCQOEPPDLKOKA-CESRCIKLEYDP--RCV-----YDTGATNORHPGE 53
DB 527 CQOQ-----QAPEDARDEHFLFSSCLEVELPQERCACGRLYQYRGTGTPAA-QAHPPGE 579

```

```

QY 54 RTRG 57
DB 580 AGCG 583

```

```

RESULT 12
ID 4CL1_SOLTU
ID 4CL1_SOLTU
AC P31684;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE 4-COMMUNARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL).
GN 4CL-1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Solanum.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-91217100; PubMed-2022667;
RA Becker-Andre M., Schulze-Lefert P., Hahlbrock K.;
RT "Structural comparison, modes of expression and putative cis-acting
RT elements of the two 4-coumarate: COA ligase genes in potato.*";
RL J. Biol. Chem. 266:8551-8559(1991).
CC -1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + PYROPHOSPHATE
CC + 4-COUMAROYL-COA.
CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOIC
CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

DN 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SEMAPHORIN 3D PRECURSOR.
GN SEMA3D.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NC
NM [1]
RP SEQUENCE FROM N.A.
RA Mead R., Graves T., Wilson C.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC
CC -1- FUNCTION: INDUCES THE COLLAPSE AND PARALYSIS OF NEURONAL GROWTH
CC CONES. COULD POTENTIALLY ACT AS REPULSIVE CUES TOWARD SPECIFIC
CC NEURONAL POPULATIONS. BINDS TO NEUROFILIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- DOMAIN: STRONG BINDING TO NEUROFILIN IS MEDIATED BY THE CARBOXY
CC THIRD OF THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC004957; AAC83081.1; ALT_SEQ.
DR INTERPRO; IPR001627; -.
DR INTERPRO; IPR001899; -.
DR INTERPRO; IPR003006; -.
DR PFM; PF01403; Sema; 2.
DR PFM; PF00047; Ig; 1.
KM Signal: Immunoglobulin domain; Multigene family; Neurogenesis;
KW Depositional protein; Glycoprotein.
FT SIGNAL 1
FT CHAIN ?
FT DOMAIN ?
FT DOMAIN 257 777
FT DOMAIN 658 738 SEMAPHORIN 3D.
FT DOMAIN 739 773 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 665 731 ARG/LYS-RICH (BASIC).
FT CARBOHYD 139 731 BY SIMILARITY.
FT CARBOHYD 607 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 724 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 724 724 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 777 AA; 89651 MW; 3F7BDD7AF50F53BD CRC64;

Query Match 17.6%; Score 60.5; DB 1; Length 777;
Best Local Similarity 34.0%; Pred. No. 16;
Matches 18; Conservative 4; Mismatches 24; Indels 7; Gaps 2;

OY 7 ORCLOSQGEHDDIKAKACESRCIKLEDPKCVYDITGATNGRRHPPGEGTRRQ 59
DB 528 QLSLHRC-----DTYGRACADCC--LARDPYCAMDGNACSRYPATSKRRARQ 573

```



```

FT DOMAIN 2536 2631 IG-LIKE C2-TYPE DOMAIN 11.
FT DOMAIN 2632 2728 IG-LIKE C2-TYPE DOMAIN 12.
FT DOMAIN 2729 2828 IG-LIKE C2-TYPE DOMAIN 13.
FT DOMAIN 2829 2926 IG-LIKE C2-TYPE DOMAIN 14.
FT DOMAIN 2927 3023 IG-LIKE C2-TYPE DOMAIN 15.
FT DOMAIN 3024 3114 IG-LIKE C2-TYPE DOMAIN 16.
FT DOMAIN 3115 3213 IG-LIKE C2-TYPE DOMAIN 17.
FT DOMAIN 3214 3300 IG-LIKE C2-TYPE DOMAIN 18.
FT DOMAIN 3301 3401 IG-LIKE C2-TYPE DOMAIN 19.
FT DOMAIN 3402 3490 IG-LIKE C2-TYPE DOMAIN 20.
FT DOMAIN 3491 3576 IG-LIKE C2-TYPE DOMAIN 21.
FT DOMAIN 3577 3671 IG-LIKE C2-TYPE DOMAIN 22.
FT DOMAIN 3671 3701 LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT DOMAIN 3701 3847 EGF-LIKE 1.
FT DOMAIN 3846 3883 EGF-LIKE 2.
FT DOMAIN 3886 3924 LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT DOMAIN 3924 3966 EGF-LIKE 3.
FT DOMAIN 3966 4104 EGF-LIKE 4.
FT DOMAIN 4106 4143 LAMININ G-LIKE 3 (GLOBULAR DOMAIN V C).
FT DOMAIN 4145 4178 HEPARAN SULFATE (POTENTIAL).
FT DOMAIN 4243 4391 HEPARAN SULFATE (POTENTIAL).
FT SITE 65 67 MEDIATES MOTOR NEURON ATTACHMENT
FT SITE 71 73 (POTENTIAL).
FT SITE 76 78 BY SIMILARITY.
FT SITE 4151 4153 BY SIMILARITY.
FT SITE 4301 4303 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 332 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 766 775 BY SIMILARITY.
FT DISULFID 782 782 BY SIMILARITY.
FT DISULFID 785 794 BY SIMILARITY.

Query Match 17.68; Score 60.5; DB 1; Length 4393;
Best Local Similarity 32.28; Pred. No. 87;
Matches 19; Conservative 4; Mismatches 33; Indels 3; Caps 2;
OY 4 PCAQRCLQSCQEPDDLKQKACESRCKLEDPKCVDTGATN--QRNPPGERTRGROP 60
Db 1137 PSCDDCDGTITPSPGLYGTCE-RCSCHGHSACEPPTGACGCGNHTEGPRCEGCGP 1194

```

Search completed: March 1, 2001, 16:17:00
Job time: 440 sec

